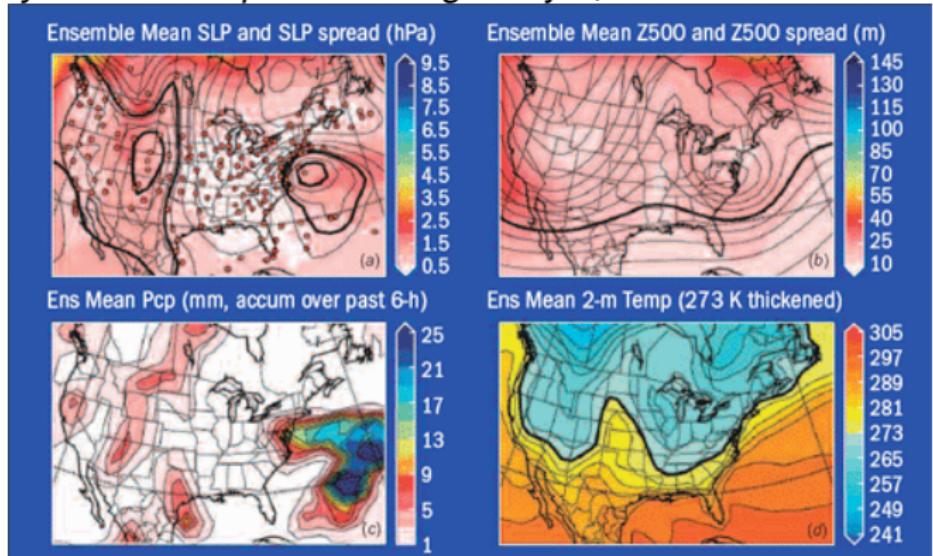




Deploying Server-side File System Monitoring at NERSC

Cray Users Group Proceedings May 7, 2009



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References

Andrew Uselton
National Energy Research Scientific Computing Center
Lawrence Berkeley National Lab

Contents

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



1 The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

2 Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

3 A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References

The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or
Intervals

Data Mining for Average
and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual
Distribution

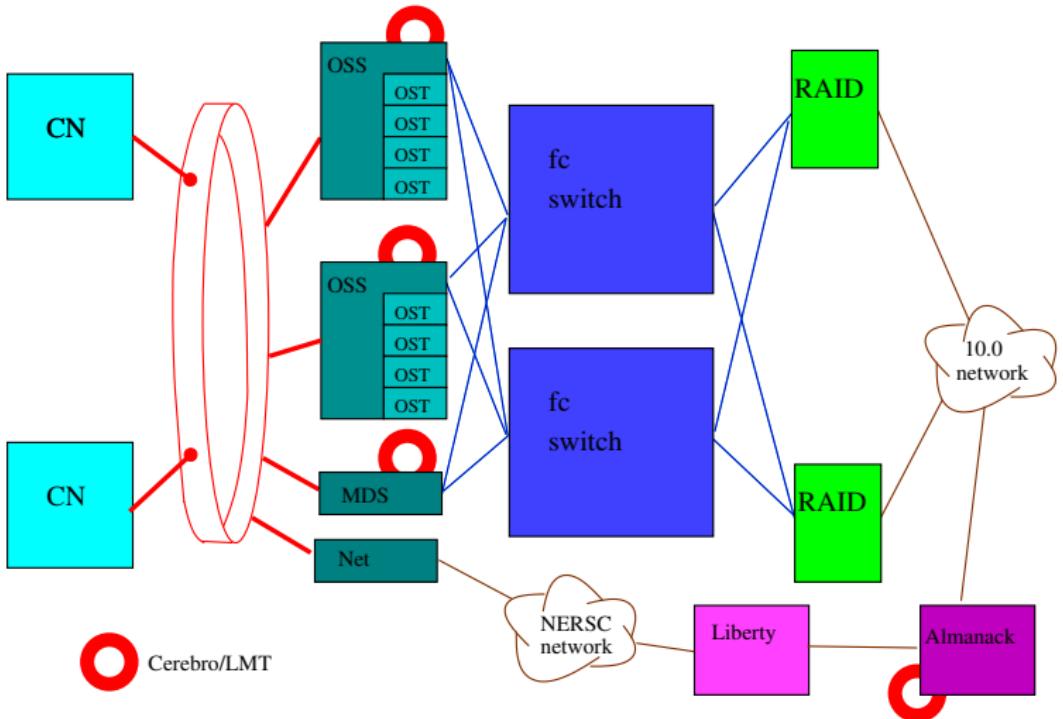
Late Breaking News

Acknowledgements and
References

Monitoring the I/O Subsystem

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

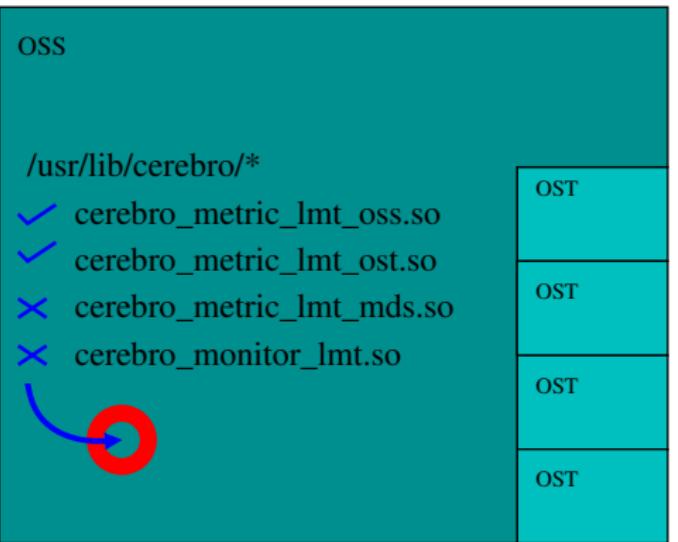
A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

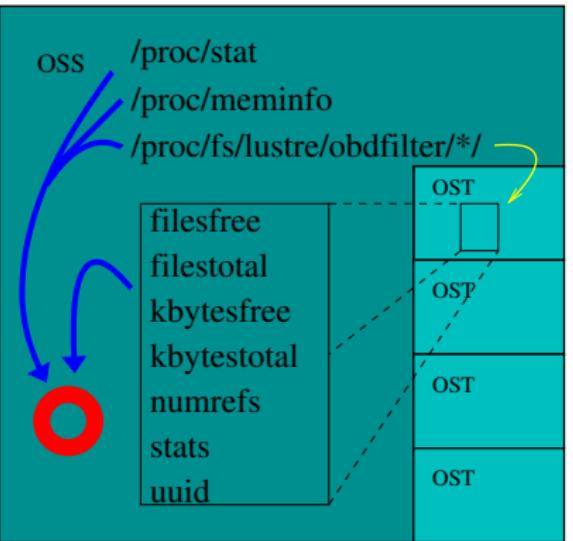
A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References



[The Franklin Cray XT4](#)

[Cerebro](#)

[The Lustre Monitoring Tool](#)

[The Lustre Dashboard](#)

Data Analysis

[Monitoring Specific Tests or Intervals](#)

[Data Mining for Average and Aggregate Behavior](#)

A Simple Model for I/O

[Poisson Distributions](#)

[Franklin's Actual Distribution](#)

[Late Breaking News](#)

[Acknowledgements and References](#)

An OSS Tuple

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



Cerebro Protocol Version
Host Name
CPU Utilization
Memory Utilization

1.0;nid04187;4.990020;39.303989

The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or
Intervals

Data Mining for Average
and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual
Distribution

Late Breaking News

Acknowledgements and
References

OST Data Values

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



Cerebro Protocol Version
Host Name
UUID
Bytes Read
Bytes Written
Kbytes Free
Kbytes Used
Inodes Free
Inodes Used

The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or
Intervals

Data Mining for Average
and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual
Distribution

Late Breaking News

Acknowledgements and
References

MDS Operations

mysql> select * from OPERATION_INFO;

OPERATION_NAME	UNITS	OPERATION_NAME	UNITS
req_waittime	usec	mds_getattr_lock	usec
req_qdepth	reqs	mds_close	usec
req_active	reqs	mds_reint	usec
reqbuf_avail	bufs	mds_readpage	usec
ost_reply	usec	mds_connect	usec
ost_getattr	usec	mds_disconnect	usec
ost_setattr	usec	mds_getstatus	usec
ost_read	bytes	mds_stats	usec
ost_write	bytes	mds_pin	usec
ost_create	usec	mds_unpin	usec
ost_destroy	usec	mds_sync	usec
ost_get_info	usec	mds_done_writing	usec
ost_connect	usec	mds_set_info	usec
ost_disconnect	usec	mds_quotacheck	usec
ost_punch	usec	mds_quotactl	usec
ost_open	usec	mds_getxattr	usec
ost_close	usec	mds_setxattr	usec
ost_stats	usec	ldlm_enqueue	usec
...			

The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or
Intervals

Data Mining for Average
and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual
Distribution

Late Breaking News

Acknowledgements and
References



The Lustre Dashboard

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton

LWatch-lustre

File Configure

nid04116_mds nid04140_mds nid00032_mds

%CPU	%KB	%Inodes
0.70	0.56	6.08
Operation	Samples	Sample /Sec
mds_close	0	0.00
mds_connect	0	0.00
mds_disconnect	0	0.00
mds_done_writing	0	0.00
mds_getattr	0	0.00
mds_getattr_lock	0	0.00
mds_getstatus	0	0.00
mds_getxattr	0	0.00
mds_pin	0	0.00
mds_quotacheck	0	0.00
mds_quotactl	0	0.00
mds_readpage	0	0.00
mds_reint	70	14.00
mds_set_info	0	0.00
mds_setxattr	0	0.00
mds_stats	0	0.00
mds_sync	0	0.00
mds_umpin	0	0.00
obd_ping	0	0.00
ALL	0	0.00

OST	Read Rate	Write Rate	%CF
ost16	0.10	2.00	****
ost8	0.40	0.31	****
ost18	0.10	0.80	****
ost5	0.23	0.21	****
ost15	0.15	1.40	****
ost7	0.35	0.95	****
ost17	0.50	2.53	****
ost9	0.25	1.72	****
ost19	0.40	0.80	****
ost46	1.70	109.44	****
ost56	0.58	0.00	****
ost48	1.05	0.40	****
ost58	0.50	0.40	****
ost45	2.06	126.70	****
ost55	0.20	0.88	****
ost47	1.09	81.44	****
ost57	0.20	0.60	****
ost49	0.30	0.41	****
ost59	0.05	0.48	****
AGGREGATE	126.49	701.26	****
MAXIMUM	26.25	126.70	****
MINIMUM	0.00	0.00	****
AVERAGE	1.58	8.77	0.01

OSS	Read Rate	Write Rate
nid0024	1.25	0.72
nid0027	0.40	3.70
nid0032	27.10	2.66
nid0035	1.31	4.90
nid0040	27.92	105.3
nid0043	2.35	4.43
nid04124	25.45	2.87
nid04127	0.55	2.25
nid04132	19.35	0.79
nid04135	1.91	4.12
nid00504	3.77	166.3
nid00507	1.20	2.78
nid00512	2.75	57.03
nid00515	1.50	1.65
nid04596	2.87	183.8
nid04599	0.65	4.94
nid04604	2.49	138.6
nid04607	1.15	5.51
nid04612	1.35	3.98
nid04615	0.80	2.78
AGGREGATE	126.14	699.3
MAXIMUM	27.92	183.8
MINIMUM	0.40	0.72
AVERAGE	6.31	34.93

The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

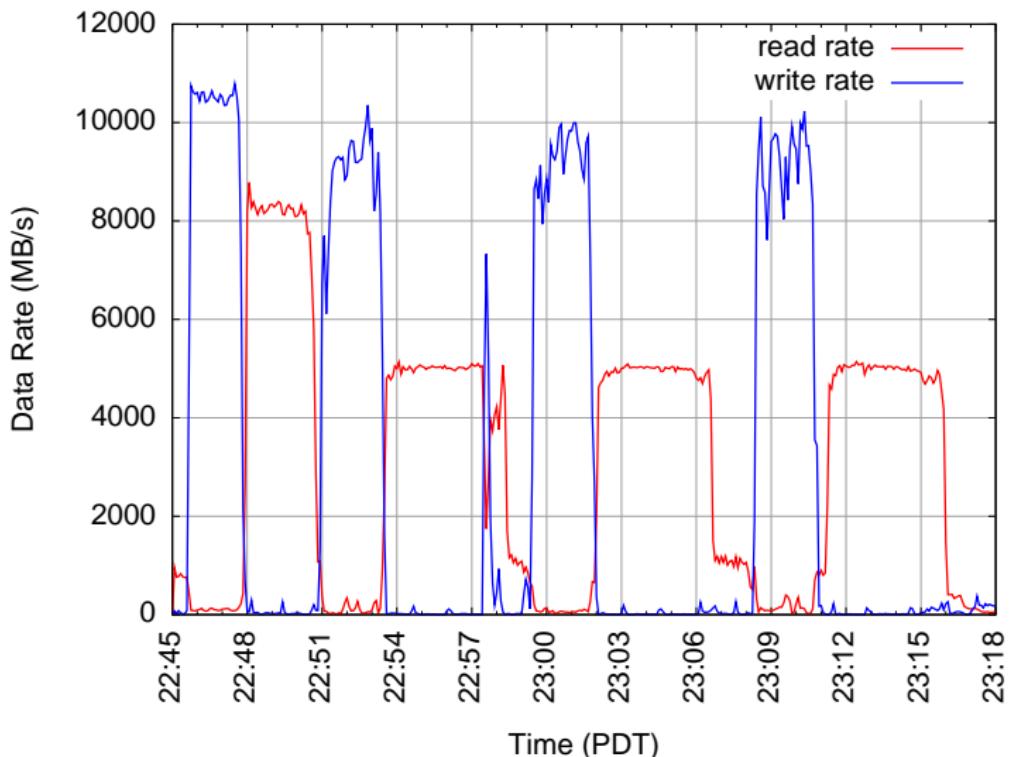
Late Breaking News

Acknowledgements and References

Four IOR Tests



Aggregate OST rates from 2008-07-28 22:45:00



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

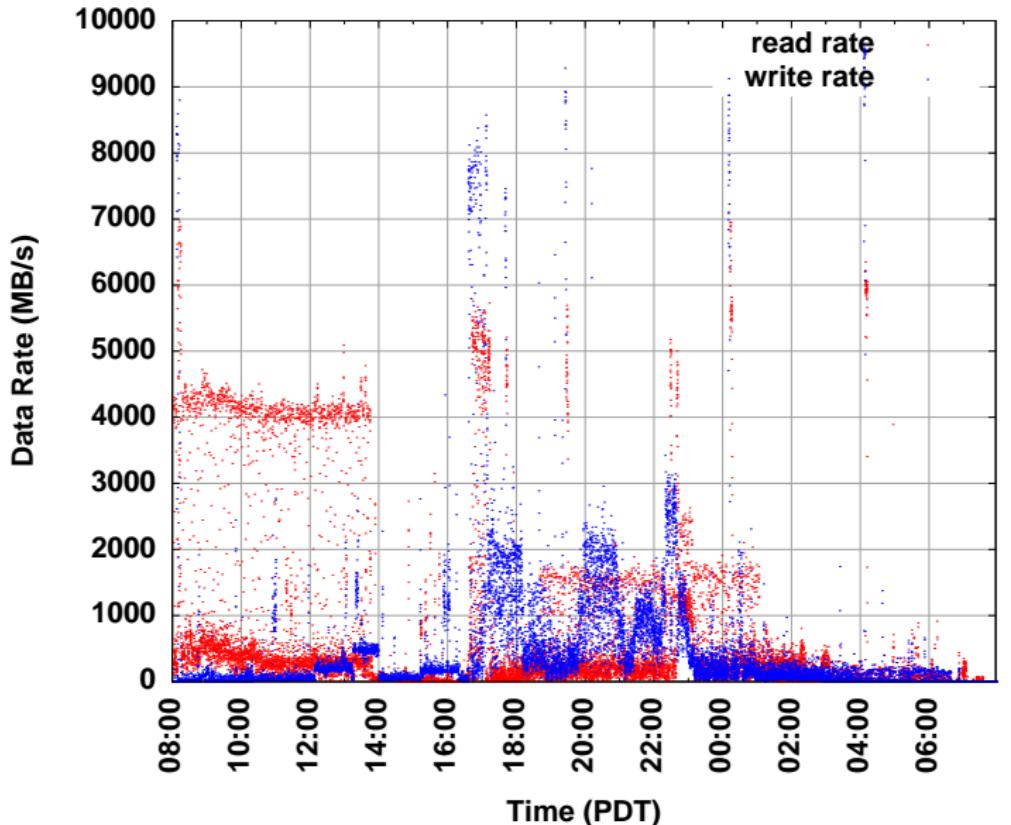
Late Breaking News

Acknowledgements and References

24 Hours of LMT Data

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

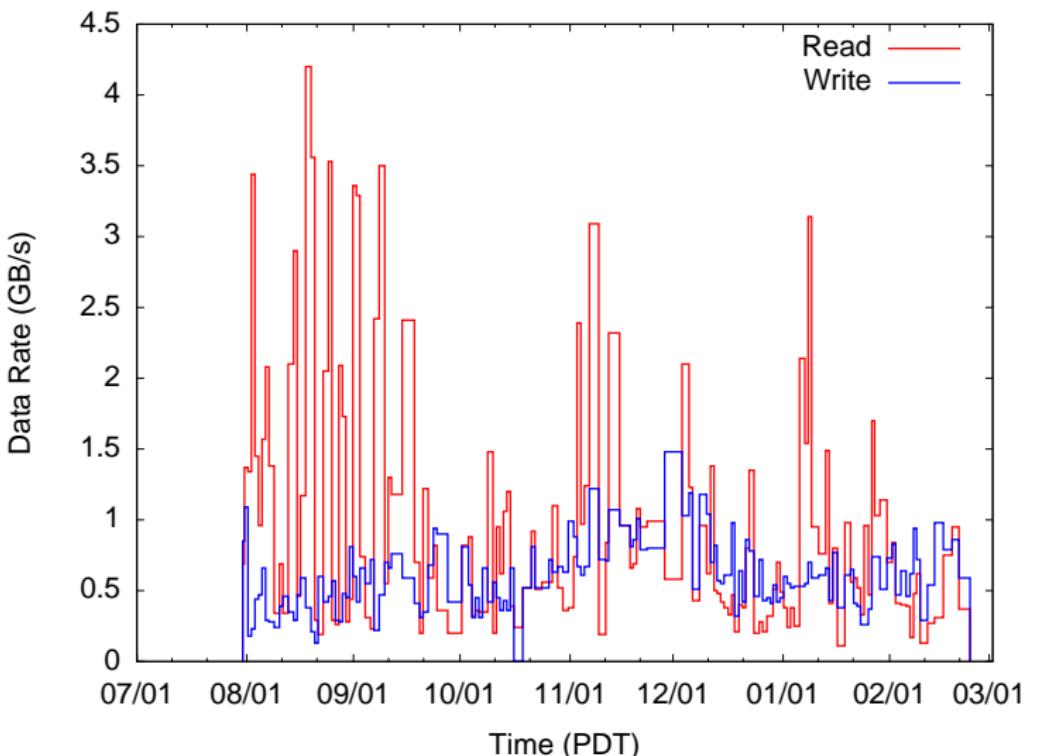
Late Breaking News

Acknowledgements and References

Daily Averages



Average daily rates



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

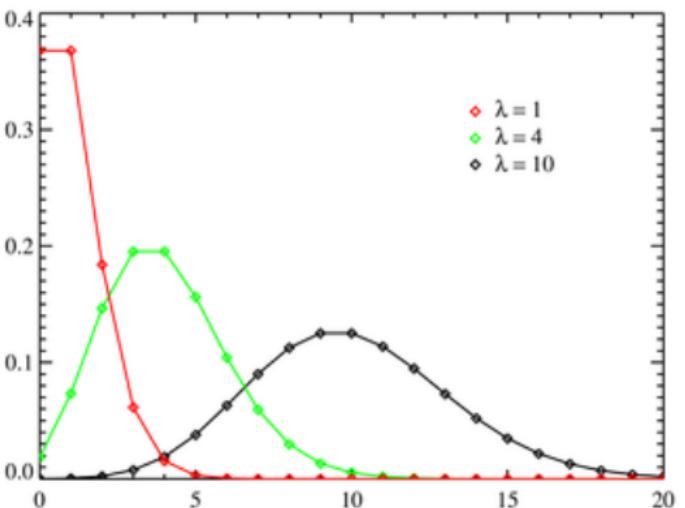
A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References



$$f_\lambda(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$





The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

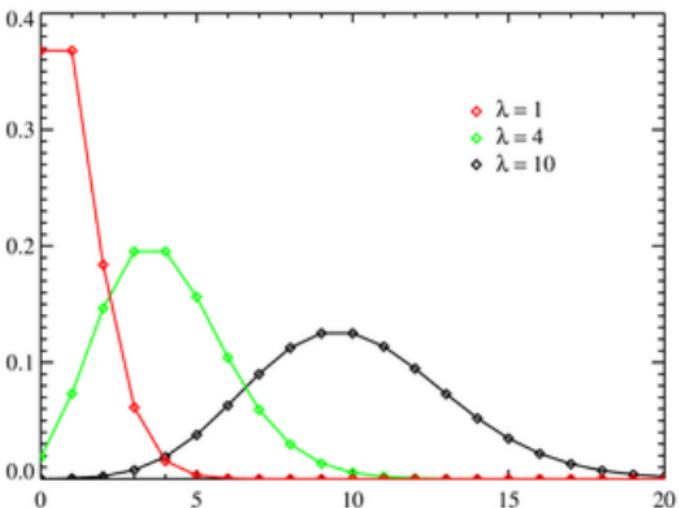
A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References



$$C(m) = N \times f_{\lambda}(\text{int}(m/M))$$



The Franklin Cray XT4

Cerebro
The Lustre Monitoring Tool
The Lustre Dashboard

Data Analysis

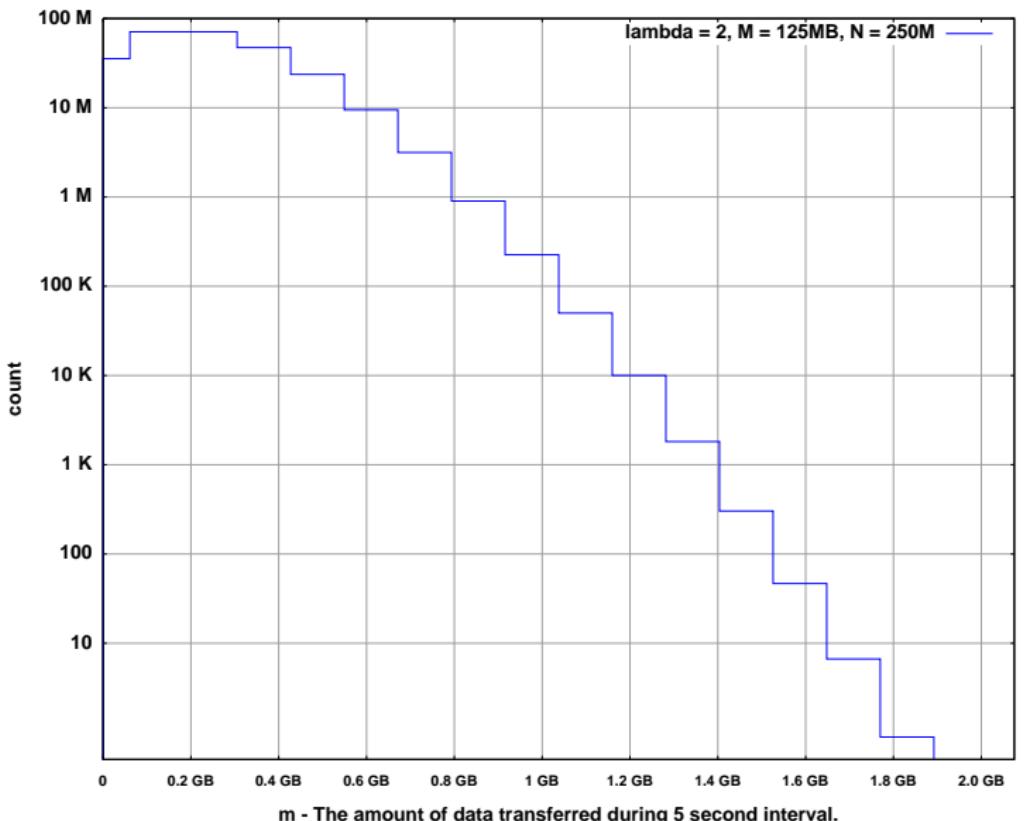
Monitoring Specific Tests or Intervals
Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions
Franklin's Actual Distribution
Late Breaking News
Acknowledgements and References

Poisson Distribution: $\lambda = 2$

Poisson distribution





The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

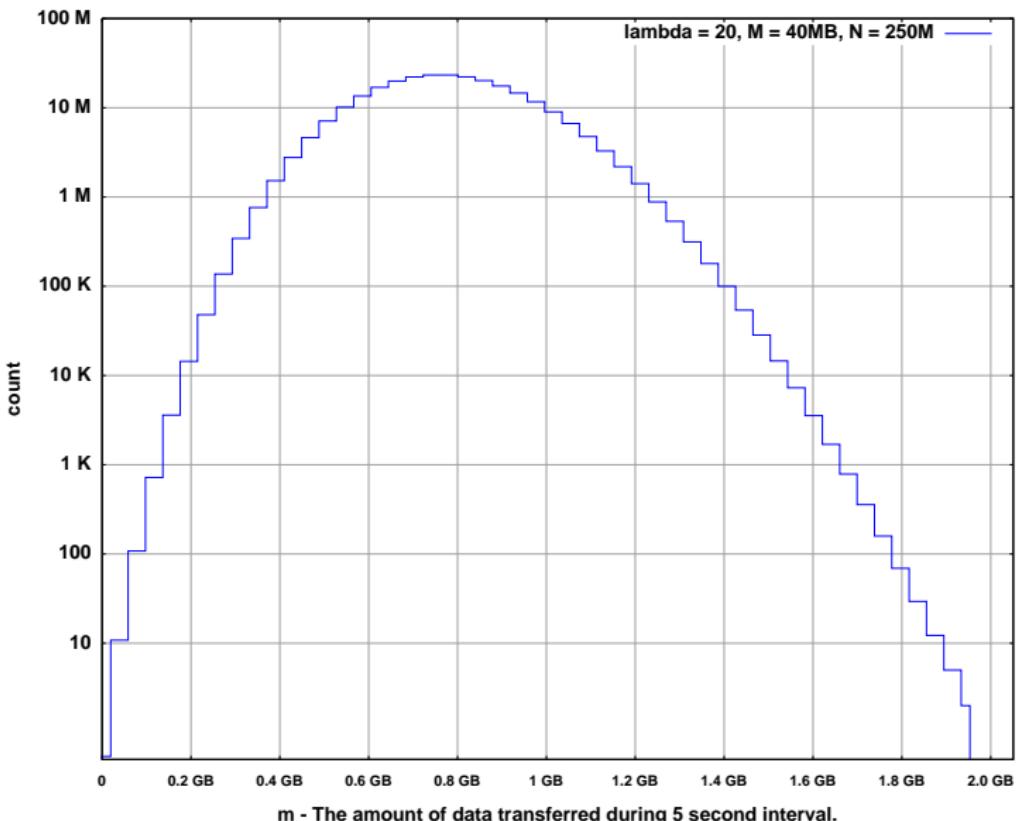
Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References

Poisson Distribution: $\lambda = 20$

Poisson distribution



250 M LMT Observations

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



The Franklin Cray XT4

Cerebro
The Lustre Monitoring Tool
The Lustre Dashboard

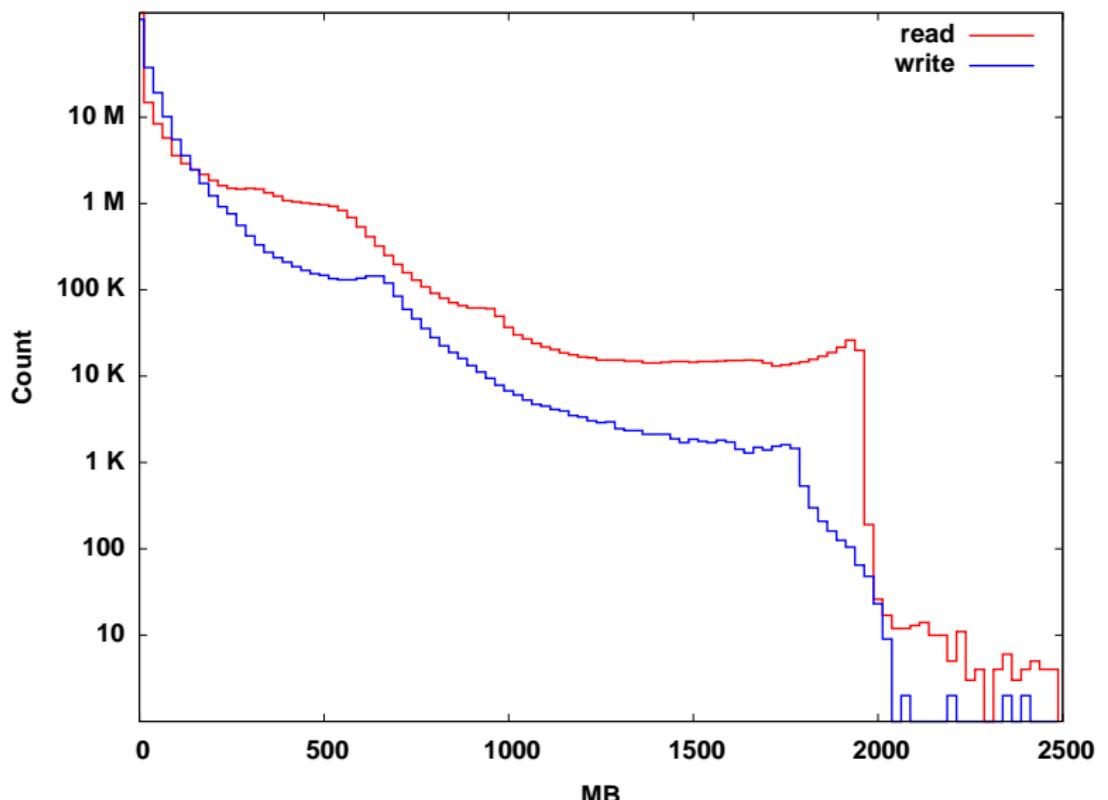
Data Analysis

Monitoring Specific Tests or Intervals
Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions
Franklin's Actual Distribution
Late Breaking News
Acknowledgements and References

Distribution of LMT observed rates



Two weeks of recent observations



The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

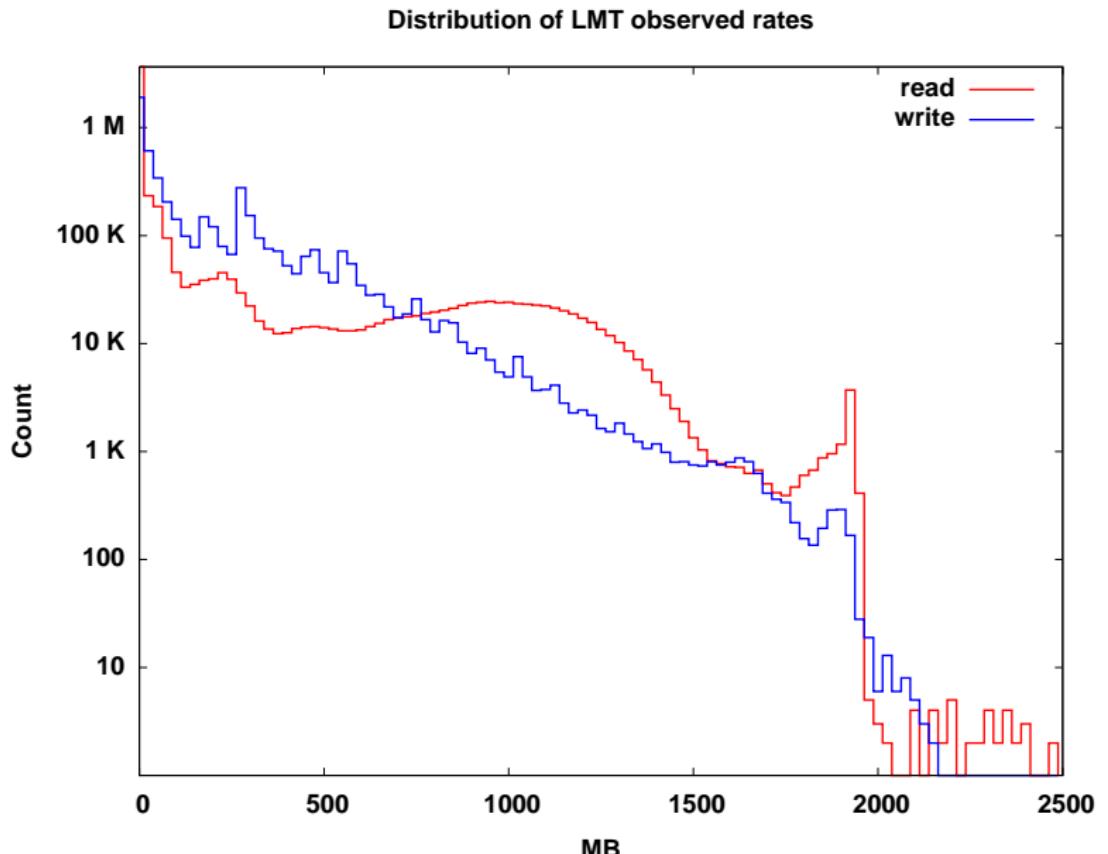
A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References



I would like to acknowledge and thank:

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



Al Chu The author of Cerebro.

Herb Wartens The author of the Lustre Monitoring Tool plug-ins.

Both work at Lawrence Livermore National Lab, which supported the development of these tools. Both were very generous with their time as I deployed the software on Franklin.

The Franklin Cray XT4

Cerebro

The Lustre Monitoring Tool

The Lustre Dashboard

Data Analysis

Monitoring Specific Tests or Intervals

Data Mining for Average and Aggregate Behavior

A Simple Model for I/O

Poisson Distributions

Franklin's Actual Distribution

Late Breaking News

Acknowledgements and References

The software is available from:

Deploying
Server-side File
System Monitoring at
NERSC

Andrew Uselton



Both applications are open source and available from Sourceforge.

Cerebro <http://sourceforge.net/projects/cerebro>

LMT <http://sourceforge.net/projects/lmt/>

If you would like hints and encouragement with getting this software deployed, contact me:

Andrew Uselton (acuselton@lbl.gov)

If you get results from your deployment that you would like to share, please do so.

[The Franklin Cray XT4](#)

[Cerebro](#)

[The Lustre Monitoring Tool](#)

[The Lustre Dashboard](#)

[Data Analysis](#)

[Monitoring Specific Tests or Intervals](#)

[Data Mining for Average and Aggregate Behavior](#)

[A Simple Model for I/O](#)

[Poisson Distributions](#)

[Franklin's Actual Distribution](#)

[Late Breaking News](#)

[Acknowledgements and References](#)